Available online at www.sciencedirect.com

SCIENCE DIRECT®

Economics Letters 84 (2004) 335–340

economics letters

www.elsevier.com/locate/econbase

Pseudo maximum likelihood estimation of structural models involving fixed-point problems

Victor Aguirregabiria*

Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA

Received 3 September 2003; received in revised form 24 February 2004; accepted 9 March 2004
Available online 20 May 2004

Abstract

This paper deals with the estimation of structural econometric models where the probability distribution of endogenous variables is implicitly defined as an equilibrium of a fixed-point problem. It proposes a pseudo maximum likelihood (PML) procedure and studies its asymptotic properties.

Keywords: Fixed-points; Pseudo maximum likelihood estimation; Recursive methods

JEL classification: C13; C63

1. Introduction

This paper deals with the estimation of structural econometric models where the distribution of endogenous variables is implicitly defined as a solution of a fixed-point problem. This structure appears in Markov discrete decision processes (Rust, 1994), auction models (Guerre et al., 2000), empirical games of incomplete information (Seim, 2002), and discrete models with social interactions (Brock and Durlauf, 2001). This paper proposes a recursive pseudo maximum likelihood (PML) procedure for the estimation of this class of models. There are two main reasons why this method is of interest. First, it avoids the problem of indeterminacy associated with maximum likelihood estimation of models with multiple equilibria. And second, the procedure avoids the repeated solution of the fixed-point problem. In models where the dimension of the fixed-point is large, this can result in significant computational gains relative to maximum likelihood.

* Tel.: +1-617-353-9583; fax: +1-617-353-4449.
E-mail address: vagguire@bu.edu (V. Aguirregabiria).

0165-1765/$ - see front matter
doi:10.1016/j.econlet.2004.03.002

It is important to note that this paper considers a model where endogenous and exogenous variables are discrete. Though the method can be extended to the case of continuous variables, the proofs of consistency and asymptotic normality of the estimator are different, and additional conditions are needed to guarantee a parametric rate of convergence for the estimator of the structural parameters.

2. Econometric model

Let $y \in Y$ be a vector of discrete random variables, where Y is a discrete and finite set, and let P^0 be the true probability distribution of y.[1] The structural model is a parametric family of probability distributions $P(\theta)$, where $\theta \in \Theta$ is a finite vector of parameters and Θ is a compact set. The model does not provide a closed form analytical expression for $P(\theta)$. Instead, this distribution is implicitly defined as a fixed-point of a mapping in probability space:

$$P(\theta) = \Psi(P(\theta), \theta) \tag{1}$$

where $\Psi(.|P,\theta)$ is a mapping from $\Im \times \Theta$ into $\Im$, where $\Im$ is the space of probability distributions of y. In some models, and for some values of θ, the mapping $\Psi(., \theta)$ can have more than one fixed-point. If that is the case, the model does not provide a unique probability distribution of y.

Example (A model of market entry). Consider a retail industry with many independent local markets. N firms are the potential entrants in each local market. Let y_{it} be the indicator of the event "firm i operates in market t," define $y_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})$, and let x_t be the size of market t. Profits of firm i in market t are:

$$\prod_{it} = \begin{cases} 0 & \text{if } y_{it} = 0 \\ \theta_{0i} + \theta_1 x_t - \theta_2 \log\left(1 + \sum_{j \neq i} y_{jt}\right) - \varepsilon_{it} & \text{if } y_{it} = 1 \end{cases} \tag{2}$$

where ε_{it} is private information of firm i and it is independently distributed over firms and over markets with distribution function F. The vector of structural parameters is $\theta = \{\theta_{01}, \ldots, \theta_{0N}, \theta_1, \theta_2\}$. By the independence of ε_{it} across firms, the joint probability $\Pr(y_t | x_t, \theta)$ can be described in terms of the set of individual entry probabilities $P(x_t, \theta) \equiv \{P_i(x_t, \theta): i = 1, 2, \ldots, N\}$. It is possible to show (see Aguirregabiria and Mira, 2003) that the equilibrium probabilities of entry are implicitly defined as the solution of the system:

$$P_i(x_i, \theta) = F(\theta_{0i} + \theta_1 x_t - \theta_2 H_i(P[x_t, \theta])) \tag{3}$$

[1] For notational simplicity we omit exogenous explanatory variables. However, all the results in this paper apply also to the case in which P^0 is a conditional probability distribution $\{P^0(y|x):(y,x) \in Y \times X\}$.

where $H_i(P)$ is the expected value of log $(1+\Sigma_{j\neq i}y_j)$ conditional on the information of firm i, and under the condition that the other firms behave according to their entry probabilities in P. That is,

$$H_i(P) = \sum_{y-i}\left(\prod_{j\neq i} P_j^{y_j}[1-P_j]^{1-y_j}\right)\log\left(1+\sum_{j\neq i} y_j\right) \tag{4}$$

and Σ_{y-i} represents the sum over all the possible actions of firms other than i. In this example, the fixed-point mapping is $\Psi(P, x_t, \theta)=\{F(\theta_{0i}+\theta_1 x_t-\theta_2 H_i(P))\colon i=1,2,\ldots,N\}$.

This example has two features that make PML estimation particularly useful. First, in general $\Psi(., x_t, \theta)$ does not have a unique fixed-point. And second, when the number of firms is relatively large, the evaluation of Ψ for different values of θ and fixed P is much cheaper than the evaluation of Ψ for different values of P and fixed θ. This is because the main computational cost comes from the sum Σ_{y-i}, and this sum should be recalculated only when we change P but not when we change θ.

3. Pseudo maximum likelihood estimators

The problem is to estimate the vector of structural parameters θ^0 given a random sample $\{y_t\colon t=1,2,\ldots,T\}$ from the population P^0. Let $\hat{P}_T^0=\hat{P}_T^0(y)\colon y\in Y\}$ be the nonparametric frequency estimator of P^0, i.e., $\hat{P}_T^0(y)=T^{-1}\Sigma_{t=1}^T I\{y_t=y\}$, where $I\{.\}$ is the indicator function. For $K\geq 1$, the K-stage PML estimator is defined as:

$$\hat{\theta}_T^K = \arg\max_{\theta\in\Theta}\sum_{t=1}^{T}\ln\Psi(y_t \mid \hat{P}_T^{K-1},\theta) \tag{5}$$

where the sequence of probability distributions $\{\hat{P}_T^K\colon K\geq 1\}$ are constructed recursively as:

$$\hat{P}_T^K = \Psi(\hat{P}_T^{K-1},\hat{\theta}_T^K). \tag{6}$$

The one-stage estimator of θ^0 maximizes the pseudo likelihood $\Sigma_{t=1}^T \ln\Psi(y_t|\hat{P}_T^0,\theta)$. Given $\hat{P}_T^0$ and $\hat{\theta}_T^1$ we obtain a new estimate of P^0 by iterating in the fixed-point mapping, i.e., $\hat{P}_T^1=\Psi(\hat{P}_T^0, \hat{\theta}_T^1)$. Then, $\hat{\theta}_T^2$ maximizes the pseudo likelihood $\Sigma_{t=1}^T \ln\Psi(y_t|\hat{P}_T^1,\theta)$, and so on.

An alternative procedure consists in calculating one-stage PML estimator and then apply one Newton iteration for the maximization of the likelihood function. There are several reasons why PML iterations may be preferred. First, a Newton iteration requires the computation of the Jacobian matrix $\partial\Psi/\partial P'$, and this can be computationally much more expensive than the successive iterations in the PML procedure. Second, in models with multiple equilibria the gradient of the likelihood function is not well defined, but the gradient of the pseudo likelihood is always well defined. And third, when the initial frequency estimator of P^0 is very imprecise, the Newton iteration estimator can perform very poorly in finite samples. Though, one can apply successive Newton iterations in the likelihood function, this can be computationally very costly.

Example. Let $\{y_t,x_t:t=1,2,\ldots,T\}$ be a sample of firms' entry decisions and market sizes from T independent local markets. For simplicity, suppose that our measure of market size is discrete. Let $\hat{P}_i^0(x)$ be the frequency estimator $(\Sigma_{t=1}^{T} y_{it}I\{x_t=x\})/(\Sigma_{t=1}^{T} I\{x_t=x\})$, and let $\hat{P}^0(x)$ be the vector $\hat{P}_i^0(x):i=1,2,\ldots,N\}$. Given these frequency estimates, the one-stage estimator maximizes the pseudo likelihood function,

$$\sum_{t=1}^{T}\sum_{i=1}^{N} y_{it}\ln F(\theta_{0i}+\theta_1 x_t-\theta_2 H_i(\hat{P}^0(x_t)))+(1-y_{it})\ln[1-F(\theta_{0i}+\theta_1 x_t-\theta_2 H_i(\hat{P}^0(x_t)))]. \tag{7}$$

When F is the *cdf* of a standard normal (logistic) random variable, this is just the likelihood of a Probit (Logit) model. Given this one-stage estimator we can get new estimates of firms' entry probabilities as: $\hat{P}_i^1(x_t)=F(\hat{\theta}_{0i}^1+\theta_1^1 x_t-\theta_2^1 H_i(\hat{P}^0(x_t)))$. Using these probabilities we can construct new values $H_i(\hat{P}^1(x_t))$, obtain a two-stage estimator, and so on.

Proposition 1 shows that the PMLEs are consistency and asymptotically normal under standard regularity conditions, and it provides a recursive expression for the sequence of asymptotic variance matrices. Proposition 2 presents a sufficient condition for the asymptotic efficiency of these estimators.

Proposition 1. *Let $\{y_t:t=1,2,\ldots,T\}$ be a random sample of y, and let $\hat{P}_T^0$ be the frequency estimator of P^0. Assume that: (a) Ψ is twice continuously differentiable in P and θ, and for any $(y,P,\theta)\in Y\times\Im\times\Theta$ the probability Ψ $(y|P,\theta)$ is strictly greater than zero; (b) $\{P(\theta):\theta\in\Theta\}$ is a correctly specified model, i.e., there is a value $\theta^0\in\Theta$ such that $P^0=\Psi(P^0,\theta^0)$; (c) Θ is a compact set; and (d) (d) θ^0 uniquely maximizes in Θ the function $E(\ln\Psi(y|P^0,\theta))$. Under these conditions the PML estimators $\{\hat{\theta}_T^K, \hat{P}_T^K:K\geq 1\}$ are root-T consistent and asymptotically normal with asymptotic variances:*

$$\mathrm{Var}(\sqrt{T}(\hat{P}_T^K-P^0))=A_K\Sigma A'_K;\quad \mathrm{Var}(\sqrt{T}(\hat{\theta}_T^K-\theta^0))=B_K\Sigma B'_K$$

where Σ is the asymptotic variance of $\sqrt{T}(\hat{P}_T^0-P^0)$, and $\{A_K:K\geq 1\}$ and $\{B_K:K\geq 1\}$ are sequences of deterministic matrices which can be obtained recursively using the expressions: $A_K=(I-\Psi_\theta M)\Psi_P A_{K-1}+\Psi_\theta M$ and $B_K=M(I-\Psi_P A_{K-1})$, where: A_0 is the identity matrix; Ψ_θ and Ψ_P are the Jacobian matrices $\partial\Psi(P^0,\theta^0)\partial\theta'$ and $\partial\Psi(P^0,\theta^0)/\partial P'$, respectively; and M is the projection matrix $(\Psi'_\theta\,diag\{P^0\}^{-1}\Psi_\theta)^{-1}\Psi'_\theta\,diag\{P^0\}^{-1}$.

Proof. By Lemma 24.1 and Property 24.2 in Gourieroux and Monfort (1995), and an induction argument, the proof of consistency is straightforward. I derive here the asymptotic distribution. First order conditions of optimality imply that the sequence of estimators $\{\hat{\theta}_T^K, \hat{P}_T^K:K\geq 1\}$ can be obtained using the recursive expressions $\Sigma_{t=1}^{T}\partial\ln\Psi(y_t|\hat{P}_T^{K-1},\hat{\theta}_T^K)/\partial\theta=0$ and $\hat{P}_T^K=\Psi(\hat{P}_T^{K-1},\hat{\theta}_T^K)$. Since Ψ is twice continuously differentiable, we can apply the stochastic mean value theorem to these conditions between $(\hat{P}_T^{K-1}, \hat{\theta}_T^K)$ and (P^0, θ^0). By consistency of $(\hat{P}_T^{K-1}, \hat{\theta}_T^K)$, the stochastic mean value theorem implies that:

$$\sqrt{T}(\hat{\theta}_T^K - \theta^0) = \Omega_{\theta\theta}^{-1}[-\Omega_{\theta P}\sqrt{T}(\hat{P}_T^{K-1} - P^0) + (1/\sqrt{T})\sum_{t=1}^{T} \partial \ln \Psi_t^0/\partial\theta] + o_p(1) \tag{8}$$

and

$$\sqrt{T}(\hat{P}_T^K - P^0) = \Psi_P\sqrt{T}(\hat{P}_T^{K-1} - P^0) + \Psi_\theta\sqrt{T}(\hat{\theta}_T^K - \theta^0) + o_p(1) \tag{9}$$

where $\Psi_t^0 \equiv \Psi(y_t | P^0, \theta^0)$, $\Omega_{\theta\theta} \equiv E(\{\partial \ln \Psi_t^0/\partial\theta\} \{\partial \ln \Psi_t^0/\partial\theta\}')$, and $\Omega_{\theta P} \equiv E(\{\partial \ln \Psi_t^0/\partial\theta\} \{\partial \ln \Psi_t^0/\partial P\}')$. Solving (Eq. 8) into (Eq. 9) we get the following recursive expression for the sequence $\{\sqrt{T}(\hat{P}_T^K - P^0)$: $K \geq 1\}$:

$$\sqrt{T}(\hat{P}_T^K - P^0) = [\Psi_P - \Psi_\theta\Omega_{\theta\theta}^{-1}\Omega_{\theta P}]\sqrt{T}(\hat{P}_T^{K-1} - P^0) + \Psi_\theta\Omega_{\theta\theta}^{-1}(1/\sqrt{T})\sum_{t=1}^{T} \partial \ln \Psi_t^0/\partial\theta + 0_p(1). \tag{10}$$

Taking into account that $\hat{P}_T^0(y)$ is the frequency estimator $(1/T)\Sigma_{t=1}^{T} I\{y_t = y\}$, we can write $(1\sqrt{T})$ $\Sigma_{t=1}^{T} \partial \ln \Psi_t^0/\partial\theta$ as $\Sigma_{y\in Y} \partial \ln \Psi(y | P^0, \theta^0)/\partial\theta \sqrt{T}(\hat{P}_T^0(y) - P^0(y))$, or in matrix form as $\Psi'_\theta \operatorname{diag}\{P^0\}^{-1} \sqrt{T}(\hat{P}_T^0 - P^0)$. Also, the matrices $\Omega_{\theta\theta} \equiv E(\{\partial \ln \Psi_t^0/\partial\theta\} \{\partial \ln \Psi_t^0/\partial\theta\}')$ and $\Omega_{\theta P} \equiv E(\{\partial \ln \Psi_t^0/\partial\theta\} \{\partial \ln \Psi_t^0/\partial P\}')$ can be written as:

$$\Omega_{\theta\theta} = \sum_{y\in Y}\{\partial \ln \Psi(y \mid P^0, \theta^0)/\partial\theta\}\{\partial \ln \Psi(y \mid P^0, \theta^0)/\partial\theta'\}P^0(y) = \Psi'_\theta \operatorname{diag}(P^0)^{-1}\Psi_\theta \tag{11}$$

$$\Omega_{\theta P} = \sum_{y\in Y}\{\partial \ln \Psi(y \mid P^0, \theta^0)/\partial\theta\}\{\partial \ln \Psi(y \mid P^0, \theta^0)/\partial\theta'\}P^0(y) = \Psi'_\theta \operatorname{diag}(P^0)^{-1}\Psi_P. \tag{12}$$

Therefore, we have:

$$\sqrt{T}(\hat{P}_T^K - P^0) = (I - \Psi_\theta M)\Psi_P\sqrt{T}(\hat{P}_T^{K-1} - P^0) + \Psi_\theta M\sqrt{T}(\hat{P}_T^0 - P^0) + o_p(1) \tag{13}$$

where $M = \Omega_{\theta\theta}^{-1} \Psi'_\theta \operatorname{diag}(P^0)^{-1}$. Solving this difference equation backwards we get $\sqrt{T}(\hat{P}_T^K - P^0) = A_K\sqrt{T}(\hat{P}_T^0 - P^0) + o_p(1)$, where $A_0 = I$ and, for $K > 0$, $A_K = (I - \Psi_\theta M) \Psi_P A_{K-1} + \Psi_\theta M$. Solving this expression in Eq. (8), we get that $\sqrt{T}(\hat{\theta}_T^K - \theta^0) = B_K\sqrt{T}(\hat{P}_T^0 - P^0) + o_p(1)$, where $B_K = M(I - \Psi_P A_{K-1})$. Finally, by Mann–Wald Theorem, it is straightforward that $\sqrt{T}(\hat{P}_T^K - P^0) \rightarrow_d N(0, A_K \Sigma A'_K)$ and $\sqrt{T}(\hat{\theta}_T^K - \theta^0) \rightarrow_d N(0, B_K \Sigma B'_K)$, where Σ is the asymptotic variance of $\sqrt{T}(\hat{P}_T^0 - P^0)$. □

Proposition 2. *If the Jacobian matrix $\partial\Psi(P^0,\theta^0)/\partial P'$ is zero, then all the estimators in the sequence $\{\hat{\theta}_T^K : K \geq 1\}$ are asymptotically equivalent to the maximum likelihood estimator (MLE).*

Proof. First, notice that applying the implicit function theorem to $P(\theta^0) = \Psi(P(\theta^0),\theta^0)$ we have that:

$$\partial P(\theta^0)/\partial\theta' = (I - \partial\Psi(P^0,\theta^0)/\partial P')^{-1}\partial\Psi(P^0,\theta^0)/\partial\theta' . \tag{14}$$

Therefore, if $\partial\Psi(P^0,\theta^0)/\partial P' = 0$, the score and pseudo-score are equal, i.e., $\partial \ln P(y_t|\theta^0)/\partial\theta = \partial \ln\Psi(y_t|P^0,\theta^0)/\partial\theta$, and we can write the variance of the MLE as:

$$V_{\mathrm{MLE}} = \left(\sum_{y\in Y}\{\partial \ln P(y \mid \theta^0)/\partial\theta\}\{\partial \ln P(y \mid \theta^0)/\partial\theta'\}\, P^0(y))^{-1} = (\Psi'_\theta \ \mathrm{diag}(P^0)^{-1}\Psi_\theta \right)^{-1} = \Omega_{\theta\theta}^{-1}. \tag{15}$$

Second, $\partial\Psi(P^0,\theta^0)/\partial P' = 0$ implies that $A_K = \Psi_\theta M$ and $B_K = M$ for any $K \geq 1$. Therefore,

$$\mathrm{Var}(\sqrt{T}(\hat{\theta}_T^K)) = M\Sigma M' = \Omega_{\theta\theta}^{-1}\Psi'_\theta \ \mathrm{diag}(P^0)^{-1}\Sigma \ \mathrm{diag}(P^0)^{-1}\Psi_\theta\Omega_{\theta\theta}^{-1} \tag{16}$$

$\Sigma = \mathrm{diag}(P^0) - P^0P^{0\prime}$ is the asymptotic variance of the frequency estimator of P^0, and it is simple to show that $\mathrm{diag}(P^0)^{-1}\Sigma\mathrm{diag}(P^0)^{-1} = \mathrm{diag}(P^0)^{-1}$. Therefore, $\mathrm{Var}(\sqrt{T}(\hat{\theta}_T^K - \theta^0)) = \Omega_{\theta\theta}^{-1}$, which is the variance of the MLE.. The *zero Jacobian condition* holds in single-agent dynamic programming models with conditional independence of unobservables (see Aguirregabiria and Mira, 2002), but it does not hold in static or dynamic games of incomplete information. However, even when the one-stage estimator is asymptotically efficient, Montecarlo experiments show that iterating in the PML procedure can provide estimators with significantly better finite sample properties (see Aguirregabiria and Mira, 2003). □

Acknowledgements

This Paper is based upon work supported by the National Science Foundation under Grant No. 0241943.

References

Aguirregabiria, V., Mira, P., 2002. Swapping the nested fixed point algorithm: a class of estimators for Markov decision models. Econometrica 70, 1519–1543.

Aguirregabiria, V., Mira, P., 2003. Sequential estimation of dynamic discrete games, Manuscript, Department of Economics, Boston University.

Brock, W., Durlauf, S., 2001. Discrete choice with social interactions. Review of Economics Studies 68, 235–260.

Gourieroux, C., Monfort, A., 1995. Statistics and Econometric Models. Cambridge Univ. Press, Cambridge, UK.

Guerre, E., Perrigne, I., Vuong, Q., 2000. Optimal nonparametric estimation of first-price auctions. Econometrica 68, 525–574.

Rust, J., 1994. Estimation of dynamic structural models, problems and prospects: discrete decision processes. In: Sims, C. (Ed.), Advances in Econometrics, Sixth World Congress. Cambridge Univ. Press, Cambridge, UK.

Seim, K. 2002. Geographic Differentiation and Firms' Entry Decisions: The Video Retail Industry. Manuscript, Stanford University.